https://doi.org/10.17221/93/2024-CJGPB

Association analysis of the molecular characteristics and floral traits of *Iris* × *germanica*

*Feng Tian[1], Ya Tian[1], Fang Yu[1], Jinsen Qian[1], Feijian Wang[1], Xue Li[1], Tongyin Li[2], Xiaofei Zhang[3], Dazhuang Huang[1], Xiaojie Zhao[1]**

[1]*Hebei Agricultural University, Baoding, P.R. China*
[2]*Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, USA*
[3]*Hebei Academy of Forestry and Grassland Science, Shijiazhuang, P.R. China*

**Corresponding author: ylzxj@hebau.edu.cn*

Citation: Tian F., Tian Y., Yu F., Qian J.S., Wang F.J., Li X., Li T.Y., Zhang X.F., Huang D.Z., Zhao X.J. (2025): Association analysis of the molecular characteristics and floral traits of *Iris* × *germanica*. Czech J. Genet. Plant Breed., 61: 55–66.

Abstract: *Iris* × *germanica* L. (bearded iris) is a popular ornamental plant with numerous commercially important cultivars; however, little is known about the genetic diversity and population structure of the species, as limited DNA markers have been explored. In this study, 34 722 expressed sequence tag (EST)-simple sequence repeat (SSR) loci were identified from RNA sequencing data. The most abundant SSR motifs belonged to the tri-nucleotide type, of which the most common were AGG/CCT followed by AAG/CTT. Overall, 50 primer pairs derived from these EST-SSRs were randomly selected and synthesized, and 22 primer pairs with good polymorphism effects were used for the following experiment. Correlation analysis of nine floral traits showed that most floral traits had significant correlations with each other. Association analysis between SSR molecular markers and nine floral traits showed that 11 EST-SSR markers were associated with 3–6 floral traits. The cluster tree constructed by using the unweighted pair group method demonstrated that the cultivars that had the same parents or similar colour were clustered together. The genotypic relations of most cultivars were consistent with their pedigree-based relationships. The EST-SSR loci identified in this study will facilitate the exploitation of genetic resources and molecular breeding of *I.* × *germanica*.

Keywords: bearded iris; EST-SSR markers; flower; genetic diversity; RNA-Seq

The genus *Iris* includes approximately 300 perennial monocotyledonous plants that are widely cultivated for their beautiful flowers. *Iris* × *germanica* is commonly known as the bearded iris and has numerous commercially important cultivars. They are descended from complex crosses and originated from Europe (Sturtevant & Randolph 1945). Two wild species, *Iris pallida* and *Iris variegate,* were major parents of the original garden cultivars of *I.* × *germanica*, and were diploids with 12 pairs of chromosomes (Sturtevant 1961). The garden diploids were crossed with a series of wild tetraploid irises (*Iris cypriana*, *Iris mesopotamica*, etc.) and formed the modern tall bearded iris. With complex genetic backgrounds, the thousands of registered cultivars of *I.* × *germanica* are hard to distinguish using only their morphological characteristics. Currently, the breeding process of *I.* × *germanica* has been hindered because of the ambiguous germplasm resources.

Compared to morphological evaluation, molecular markers are not affected by the environment; therefore, they have been widely used by plant researchers

Supported by S&T Program of Hebei (21326317D) and Project for PhD of Hebei Agricultural University (ZD2016025).

https://doi.org/10.17221/93/2024-CJGPB

to identify cultivars and analyse hybrid origin and genetic diversity (Agarwal et al. 2008). Traditional breeding programs are based on hybridization, which is limited by the long selection cycle. Marker-assisted selection (MAS) enables breeders to quickly select cultivars with desirable traits and has great potential to accelerate crop breeding processes. The two effective approaches commonly used to develop trait-linked markers are conventional linkage analysis in controlled crosses and linkage disequilibrium (LD)-based association analysis using diverse germplasms (Brachi et al. 2010; Butiuc-Keul et al. 2022; Guo et al. 2017; Wang et al. 2020). Compared with traditional linkage mapping approaches, association mapping exploits the recombination events that occurred during evolutionary history based on existing natural populations (Mazzucato et al. 2008; Zhen et al. 2018; Liu et al. 2022).

Different types of DNA molecular markers have been developed and successfully applied in genetics and breeding activities in various plants. For instance, restriction fragment length polymorphism (RFLP) (Williams 1989), amplified fragment length polymorphism (AFLP) (Paun & Schönswetter 2012), random amplified polymorphic DNA (RAPD) (Upadhyay et al. 2010a), microsatellites or simple sequence repeats (SSRs) (Upadhyay et al. 2010b), inter-simple sequence repeats (ISSRs) (Hinge et al. 2022), and cleaved amplified polymorphic sequences (CAPS) markers (Chavhan et al. 2023) have been utilized to detect the genetic basis of variation, genetic relationships within germplasm, and for the identification of duplicate accessions.

DNA-based molecular markers are not affected by environmental conditions and growth stages; therefore, they are useful in genetic diversity analyses (Lee & Park 2017). RAPD markers were used to analyse the genetic variation of Iranian iris species (Azimi et al. 2012). ISSR markers have been shown to be a suitable approach for the evaluation of molecular diversity and phylogenetic relationships in *Iris* spp. (Attari et al. 2016). AFLP markers were used to study genetic diversity and the population structure of species or cultivars derived from *I.* × *germanica*, *I. pumila*, *I. variegata*, and *I. pallida* (Li et al. 2020). Three types of PCR markers (ISSR, mobile genetic element sequences (Inter-Retrotransposon Amplified Polymorphism and inter-Primer Binding Site), and abiotic stress response genes (LP-PCR) were used to assess the genetic polymorphism in *Iris pumila*. LP-PCR markers showed the highest level of genetic polymorphism, while ISSR markers could identify all individual plants in the population (Bublyk et al. 2021). Five chloroplast DNA (cpDNA) fragments and seven microsatellite markers were used to investigate the genetic diversity, population structure, and historical dynamics of *Iris loczyi* (Zhang et al. 2021).

Among the various kinds of molecular markers, SSRs are short tandem repeats of one to six nucleotides that are adequately dispersed throughout the coding and non-coding regions of the genome; their distribution is also highly organized and varies in different regions of the genes (Lawson & Zhang 2006). In different individual genotypes, the number of repeat units could vary due to changes in tandem arrays in the SSR motifs. SSRs have been widely used due to their characteristics of co-dominant inheritance, multi-allelic nature, extensive genome coverage, and simple detection (Varshney et al. 2005). Additionally, expressed sequence tag (EST) SSR markers can be rapidly mined at lower cost from expressed sequences and have the potential for direct gene tagging for important horticultural traits (Bouck & Vision 2010; Wu et al. 2014; Santos et al. 2018; Riangwong et al. 2020). In recent years, next-generation sequencing (NGS) technology has enabled researchers to identify a large number of SSRs based on abundant sequence information (Ronoh et al. 2018; Šarhanová et al. 2018; Zhu et al. 2021).

Currently, upgraded Illumina technology is the most widely utilized NGS platform for de novo transcriptome sequencing (RNA-Seq) for developing SSR markers (Taheri et al. 2018). For plants without a reference genome, transcriptome sequence can be used for the identification and development of SSR loci. Several ornamental plant species lack reference genomes; however, Illumina sequencing was used in studies of calla lily (*Zantedeschia rehmannii* Engl.) (Taheri et al. 2018), *Curcuma alismatifolia* (Taheri et al. 2019), Lilac (*Syringa oblata*) (Yang et al. 2020), of which 9 351, 9 933, and 10 988 SSRs, respectively, were mined. Up to 200 primer pairs were designed and synthesized according to the transcript sequence; 58 were polymorphic among 21 accessions of coloured calla lily (Wei et al. 2016). The validated and polymorphic SSR markers were used to assess the genetic diversity, population structure, and further association mapping.

Simple sequence repeat markers were developed in some species of *Iris* L., such as *Iris loczyi* (Zhang et al. 2021), *I. laevigata* (Sun et al. 2012), *I. ensata* (Xiao et al. 2012), *I. brevicaulis,* and *I. fulva* (Tang et al. 2009). However, few SSR markers have been

https://doi.org/10.17221/93/2024-CJGPB

developed in *I.* × *germanica*, because the cross-specific amplification of microsatellites may result in a high possibility of null alleles (Jiang et al. 2011). Therefore, there is a need to develop SSR markers for MAS in *I.* × *germanica*. In this study, polymorphic EST-SSRs were developed and used to construct pedigree trees and explore specific allelic variations between phenotypes and genotypes of *I.* × *germanica* cultivars.

MATERIAL AND METHODS

All plant materials (Figure S1 in Electronic Supplementary Material (ESM)) used in this research were planted in the nursery garden of Hebei Agricultural University in Baoding, China (38°49'30''N, 115°26'44''E). Most of the cultivars were obtained from Schreiner's Iris Gardens (Salem, Oregon, USA), while cv. Indian Chief and cv. White and Yellow were obtained from the Hebei Academy of Forestry and Grassland Science (Shijiazhuang, Hebei, China).

Isolation of microsatellite markers using transcriptome sequencing. For transcriptome sequencing, total RNA was extracted from leaves of two cultivars, Immortality and New moon, using the RNeasy Plant Mini Kit (Qiagen, Germany). Paired-end sequencing cDNA libraries were constructed and sequenced using the Illumina HiSeq 2500 system (Illumina Inc., CA) according to the manufacturers' instructions. The raw reads were cleaned by removing reads containing poly-N and low-quality reads. The cleaned reads were then assembled in the Trinity software using default parameters.

Perl scripts were developed to search for SSRs using the MIcroSAtellite (MISA, Thiel et al. 2003) search module. The parameters were set to detect perfect ten-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 10, 6, 5, 5, 5, and 5 repeats, respectively. After identifying EST-SSR markers, Primer 3 was used to design SSR primers using default parameters (Untergasser et al. 2012).

Primer screening and cross-species amplification. To test the amplification efficiency of the newly developed EST-SSR markers, 50 primer pairs were randomly selected and synthesized, of which 22 pairs showed polymorphism. Twenty-five commercial cultivars (Table 1, Figure 1) were used to confirm

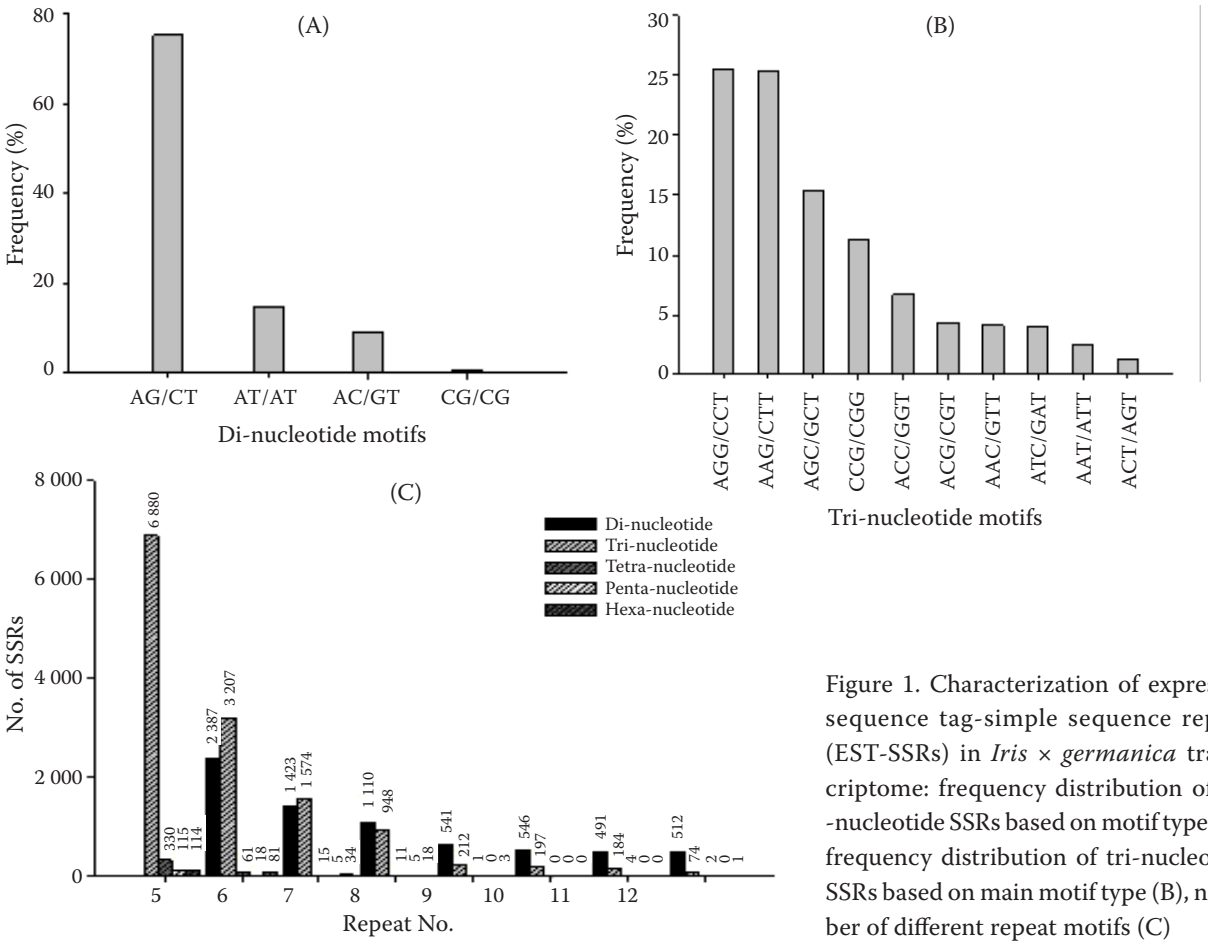

Figure 1. Characterization of expressed sequence tag-simple sequence repeat (EST-SSRs) in *Iris* × *germanica* transcriptome: frequency distribution of di-nucleotide SSRs based on motif type (A), frequency distribution of tri-nucleotide SSRs based on main motif type (B), number of different repeat motifs (C)

Czech Journal of Genetics and Plant Breeding, 61, 2025 (2): 55–66

https://doi.org/10.17221/93/2024-CJGPB

Table 1. Pedigrees, types, and breeders for 25 *Iris* × *germanica* genotypes*

Code	Cultivar name	Blooming types	Flower colour	Breeder	Year of release	Pedigree
C1	About town	once-bloomer, TB	silvery mauve lilac and red violet	Barry Blyth	1996	Bubble Up × Electrique
C2	Angelwalker	re-bloomer, IB	white	Terry Aitken	2004	Pure Allure × Frost Echo
C3	Autumn circus	re-bloomer, TB	white and violet	Ben Hager	1990	(Space Odyssey × Socialite) × Earl of Essex
C4	Autumn tryst	re-bloomer, TB	white and rosy lavender	John Weiler	1993	Lilac Stitchery × Earl of Essex
C5	Blueberry Tart	re-bloomer, SDB	blue and reddish tan	Chuck Chapman	2002	Forever Blue × What Again
C6	Celebration Song	once-bloomer, TB	apricot pink and blue lavender	Schreiners Gardens	1993	Lullaby of Spring × Frances Gaulter
C7	Chariots of Fire	once-bloomer, TB	peach pink	Terry Aitken	2000	seedling# 85 T1 × seedling# 87 T96
C8	Clarence	re-bloomer, TB	white and violet	Lloyd Zurbrigg	1990	parentage unknown
C9	Girly Girl	re-bloomer, TB	Standards and falls white ground, 1/4" purpleplicata markings	Schreiners Gardens	2013	parentage unknown
C10	Hot	re-bloomer, SDB	yellow and brown	Monty Byers	1989	Sunstrip × Frankincense
C11	I'm back	re-bloomer, TB	red-purple and purple black	Thomas Johnson	2006	Brazilian Holiday × Romantic Evening
C12	Immortality	re-bloomer, TB	white	Lloyd Zurbrigg	1982	I Do × English Cottage
C13	Indian Chief	once-bloomer, TB	red-purple	Wylie McLean Ayres	1929	Cardinal × unknown
C14	Kaligazam	once-bloomer, TB	raspberry pink	Graeme Grosvenor	2000	Anna Glitsch × Elizabeth Marrison
C15	Lenora Pearl	re-bloomer, BB	salmon pink	Hooker Nichols	1988	Natural Beauty × Bubble Up
C16	Magical Encounter	re-bloomer, TB	pink	Schreiners Gardens	1999	seedling# S 367-A × Dreamsicle
C17	Midsummer Night's Dream	re-bloomer, IB	dark purple	Lowell Baumunk	1998	Best Bet × What Again
C18	Pallida Variegata Gold	once-bloomer, TB	violet	E.H. Krelage and Sons	1900?	unknown
C19	Pallida Variegata Silver	once-bloomer, TB	violet	E.H. Krelage and Sons	1900?	unknown
C20	Party's Over	once-bloomer, TB	lavender blue and ivory	Duane Meek	2005	Color Magician × Champagne Frost
C21	Peach Jam	re-bloomer, TB	pink with streaked purple and mauve	Allan Ensminger	1989	seedling# 75-18 × seedling# 78-21
C22	Puddy Tat	once-bloomer, SDB	blue white and blue black	Paul Black	2002	Well Suited × Jazzamatazz
C23	Sugar Blues	re-bloomer, TB	wisteria blue	Lloyd Zurbrigg	1984	Victoria Falls × unknown
C24	Summer Olympics	re-bloomer, TB	greenish yellow and white	Raymond Smith	1976	((Rippling Waters × seedling# 121BR) × Winter Olympics)
C25	White and Yellow	re-bloomer, TB	white and yellow	unknown	unknown	unknown

*Information compiled from The America Iris Society (2019); genotypes are ordered alphabetically; TB – tall bearded; IB – intermediate bearded; SDB – standard dwarf bearded; BB – border bearded

https://doi.org/10.17221/93/2024-CJGPB

the cross-species transferability of the 22 selected EST-SSR markers.

Genomic DNA from healthy and young leaves was extracted using the modified CTAB method (Doyle & Doyle 1987). DNA quality was detected using 1.5% agarose gel electrophoresis. The concentration of DNA was measured using a micro spectrophotometer NanoDrop 2000 (Thermo Fisher Scientific, US) and work concentration was diluted to 200 ng/μL.

Afterwards, SSR-PCR amplification was performed in a volume of 20 μL, containing 10 ng DNA, 1.2 μL of 2.5 mmol/L dNTPs, 2 μL 10 × PCR buffer (Mg^{2+}), 1 U *Taq* DNA polymerase (Takara Biotechnology Company, Dalian, China), and 0.8 μL of 10 μmol/L forward and reverse primers. The PCR conditions were as follows: 95 °C for 1 min; 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s or 90 s; 72 °C for 10 min; and hold at 4 °C. PCR was performed using T100 Theral Cycler (Bio-Rad, USA). The amplification products were electrophoresed on 6% polyacrylamide gels (19 : 1 acrylamide:bis) using the Thermo Scientific™ Owl™ vertical electrophoresis systems (Thermo Fisher Scientific, USA).

Phenotypic data. A total of nine floral traits were evaluated in 25 *I.* × *germanica* cultivars with three repeats per genotype (Table 2). The nine floral traits included flowering date (FD), number of florets per stem (NF), length of flower stem (LS), length of fall petal (LFP), width of fall petal (WFP), length of standard petal (LSP), width of standard petal (WSP), height of floret (HF), and width of floret (WF), which were all recorded or measured at full bloom. LS, LFP, WFP, LSP, WSP, HF, and WF were measured using a measuring tape. Flowering dates were recorded when the first floret opened. Flowering dates were divided into the following five groups: 1, very early; 3, early; 5, medium; 7, late; 9, very late.

Data analyses. The EST-SSR allelic composition of each accession was recorded as dominant markers, and according to the amplicon size, putative alleles were labelled with different letters indicating the different alleles. Afterwards, a dominant genotypic matrix was formed and used for the calculation of polymorphisms of SSR markers using Power Marker version 3.51 (Yu & Buckler 2006). TASSEL with the General Linear Model (GLM) option (Liu & Muse 2005) was used to examine associations between phenotypic traits and DNA markers. Taking the Q matrix of population structure analysis as a covariable, the nine floral traits and SSR polymorphism marker data were substituted into the general linear model in TASSEL software (Bradbury et al. 2007) for regression analysis. To illustrate the genetic relationships among the cultivars, a dendrogram was constructed using the un-weighted pair group method with arithmetic averages (UPGMA) using the NTSYS-pc program (Ver. 2.2, Rohlf 2005).

RESULTS

Frequency and distribution of EST-SSRs loci in *Iris* × *germanica*. The MIcroSAtellite search module was used to search for SSRs in the sequences of 124 758 unigenes. Overall, 34 722 potential EST-SSR loci were identified from 26 739 unigenes sequences, among which 6 070 sequences contained more than

Table 2. Estimates of phenotypic correlations for nine floral traits in the association population in *Iris* × *germanica*

Traits	FD	NF	LS	LFP	WFP	LSP	WSP	HF	WF
FD	1								
NF	0.463*								
LS	0.625**	0.550**							
LFP	0.437*	0.283	0.723**						
WFP	0.483**	0.129	0.639**	0.805**					
LSP	0.617**	0.441*	0.775**	0.720**	0.504**				
WSP	0.395**	0.250	0.669**	0.814**	0.836**	0.553**			
HF	0.585**	0.479**	0.841**	0.768**	0.597**	0.816**	0.626**		
WF	0.311*	0.030	0.495**	0.681**	0.854**	0.277	0.735**	0.417**	1

FD – flowering date; NF – number of florets per stem; LS – length of flower stem; LFP – length of fall petal; WFP – width of fall petal; LSP – length of standard petal; WSP – width of standard petal; HF – height of floret; WF – width of floret; *,**$P < 0.05$, 0.01 (significance of Pearson correlation coefficients between two traits)

https://doi.org/10.17221/93/2024-CJGPB

one EST-SSR locus, indicating that 21.4% of unigenes contained SSR loci. Among the di-nucleotide motifs, the most common motifs type was AG/CT (72.3%), followed by the AT/AT motif (14.8%) (Figure 1A). The most abundant SSR motifs in the tri-nucleotide SSRs were AGG/CCT (25.4%) and AAG/CTT (25.2%), followed by AGC/GCT (15.3%) (Figure 1B). The counted repeat unit number of SSRs ranged from 5 to 12, and the top two abundant SSR types were SSRs with 5 and 6 repeats (Figure 1C).

Table 3. Descriptions of expressed sequence tag-simple sequence repeat (EST-SSR) markers used for genotyping detection of *Iris* × *germanica*

Primer name	Sequence (5'- 3')	Repeat motif	Allele ranges (bp)	Number of alleles	Shannon index
IG19F1	F: GAGAACACATCACACTC R: AGACCGTCAACGTGTAAGCC	(GAA)10	232–259	11	0.57
IG19F3	F: TGGAAACTTGAGGTGCACAA R: GGCTCCATTTCTGCTAGTCG	(GAT)10	126–141	8	0.55
IG19F5	F: CGCCAGGTGCTCGTTAATA R: ACAACAAGAAGCCTACCAGC	(CTC)7	272–290	8	0.60
IG19F6	F: ATGCTCAGATCGAGGAGGAA R: ATCTGCAGCCTACACCACCT	(ACG)24	254–278	4	0.56
IG19F10	F: AACGTGGCCAGAGTGCTTTA R: TGCTTCTGTTGTTGGGGAAT	(TCCTT)5	246–261	5	0.64
IG19F14	F: TAGGCCATCAAACGATTCCT R: TTCTGAGCAAAAGTCCGACA	(A)44	280–298	7	0.58
IG19F15	F: AGCCTATGGTGGATCAGTGC R: TGGAAAGCATGTGTACCTTCA	(AAG)13	279–306	6	0.56
IG19F17	F: AGCCTATGGTGGATCAGTGC R: TGGAAAGCATGTGTACCTTCA	(CT)25	241–263	7	0.57
IG19F22	F: TGGTGGTGATTTGTTTCTTGA R: TGATGGTGTGGTTGGTTGTT	(GAA)14	260–281	6	0.61
IG0809F4	F: AATCCGGACCTCAACCAGTG R: ACCGATCGAATACTTGGGACC	(ACC)5	223–241	5	0.52
IG0821F2	F: TCGGTTCGGTTTAAGGTCGG R: GTTCCCATGTGCATGCTGTG	(A)13	104–115	2	0.61
IG0821F4	F: AAGTTGGAGGAGGAGGTGGT R: GATAAAGGAGGAGGCGAGCG	(TGC)5	257–279	4	0.52
IG1001F12	F: TGCACTGCGACATCAACTCT R: CACCACCACCACCTCTTCTC	(GTT)10	165–186	6	0.61
IG1002F14	F: TCCCCCTTCATTTCTTCGGC R: CGGCAGCTTTTCCATCCAAC	(T)16	106–119	4	0.52
IG1111F15	F: TCCTTCCCGTCTACCTCCTG R: ACCCCCGTCTCTTCACTCTT	(TC)8	166–186	4	0.57
IG1212F17	F: CCTGCCACTCTTCACACCAA R: AGCGCCACTCCAACTAACAA	(A)15	220–231	4	0.51
IG1212F18	F: CCTCCCTCTCCTCACTCTCC R: TCGGTTCGGTTCGGTTTCAT	(A)12	178–190	2	0.52
IG1212F19	F: CCCTTCCGTTCTTCTCGTCC R: TTGCTGCCTGTGCTCTTCTT	(TGG)5	171–186	9	0.56
IG1212F20	F: TCCATCTCAGCTTCCTCGGA R: CCCACCCTATCGCCAACAAA	(GAG)5	236–251	2	0.59
IG1217F22	F: AGACCCAGATACCGCAGAA R: GGATGATGAGCCTTACCGT	(TC)9	280–287	5	0.48
IG0121F47	F: TGCCATCACTCAACACCA R: TCCCTTCCTCTTCTTCGA	(AG)7	146–160	5	0.62

https://doi.org/10.17221/93/2024-CJGPB

Phenotypic floral trait distribution and correlations. Pearson's correlation coefficients between pairs of the nine floral traits were calculated, and 31 significant correlations were shown among these traits ($P < 0.05$) (Table 2). Positively significant correlations were also found between most floral traits. The length of the flower stem (LS) had a significant positive correlation with all floral traits. The traits of the florets, such as the length of fall petal (LFP), width of fall petal (WFP), length of standard petal (LSP), width of standard petal (WSP), height of floret (HF), and width of floret (WF), were positively significantly correlated with any other two traits. The cultivars that flowered late also had taller flower stems and larger florets.

Correlation analysis between floral traits and EST-SSR molecular markers in *Iris* × *germanica*. Fifty pairs of primers derived from EST-SSR were randomly selected and synthesized, and 22 pairs of primers with good polymorphism effects were screened (Table 3). The association analysis between SSR molecular markers and nine floral traits was performed using the results of population structure analysis to avoid false positive correlation (Table 4). Using $P < 0.05$ as the benchmark, 11 markers were found to be associated with at least one floral trait. The marker IG1212F18 was associated with seven floral traits (FD, NF, LS, LFP, WSP, HF, and WF), and this marker also showed high correlation with LS and HF.

Application of SSR markers in studying the genetic relationship of *Iris germanica* cultivars. The UPGMA dendrogram was constructed for 25 iris cultivars based on the 22 EST-SSR markers. The 25 cultivars were separated into four groups using a similarity coefficient of 0.6 as the threshold (Figure 2). Cluster I contained 15 cultivars and was the largest group. Cluster II only had one cultivar. The cultivars that had the same parents were clustered together, such as Autumn Circus (C3) and Autumn Tryst (C4), Blueberry Tart (C5) and Midsummer Night's Dream (C17) (Figure 2).

Two cultivars of *Iris pallida*, Pallida variegata gold (C18) and Pallida variegata white (C19), were also clustered together (Figure 2). The cultivars that had similar colours also showed high similarity coefficients. For instance, cluster III had three pink colour cultivars; Chariots of Fire, Celebration Song, and Kaligazam clustered together. Based on the common horticultural classification, the iris cultivars were also classified as once-bloomers and re-bloomers. The cultivars in cluster IV all belong to the re-bloomer type. However, not all re-bloomers were grouped in one cluster.

DISCUSSION

DNA-based markers provide an effective way for assessing genetic diversity, as RAPD markers (Azimi

Table 4. Summary of significant marker-trait pairs from the association test results of *Iris* × *germanica* using the TASSEL program; under the condition of $P < 0.05$, 11 markers were associated with at least one floral trait

Markers	Traits								
	FD	NF	LS	LFP	WFP	LSP	WSP	HF	WF
IG1111F15			*					*	**
IG1212F18	*	*	****	**			**	****	**
IG1212F19	*			*		*			
IG1212F20			*				*		*
IG1212F22		*							
IG19F1					*				*
IG19F6		*		**	*	*			*
IG19F14	**				**				*
IG19F17						**	****		
IG19F22		*	****	**	**				
IG19F15		**							

FD – flowering date; NF – number of florets per stem; LS – length of flower stem; LFP – length of fall petal; WFP – width of fall petal; LSP – length of standard petal; WSP – width of standard petal; HF – height of floret; WF – width of floret; *,**,***$P < 0.05$, 0.01, 0.0001 (significance of association between markers and traits)

https://doi.org/10.17221/93/2024-CJGPB

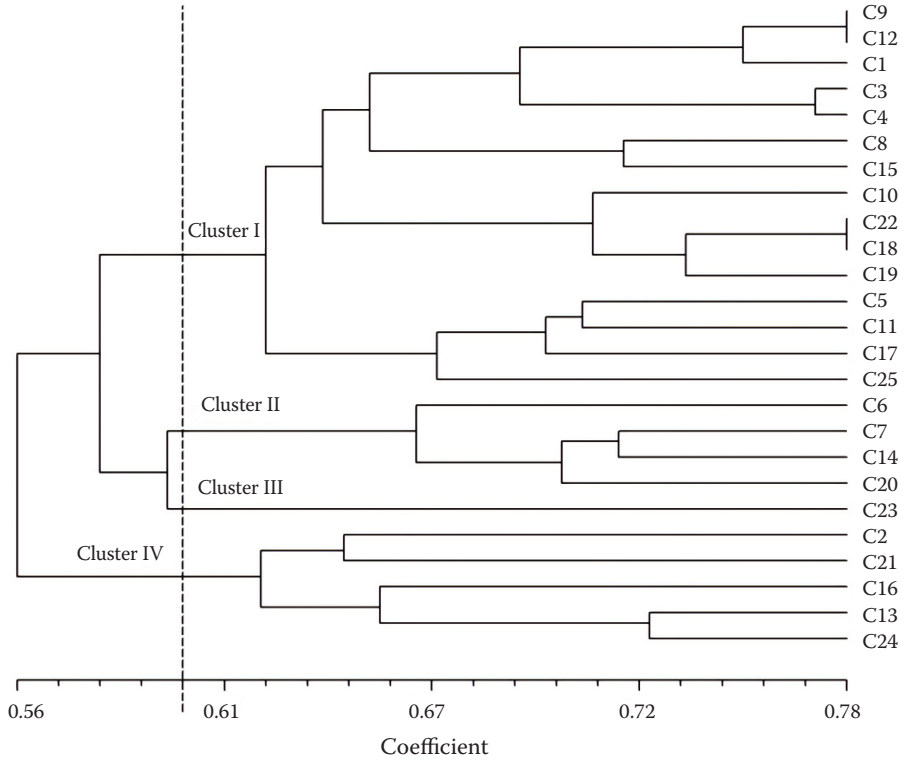

Figure 2. Dendrogram showing the genetic affiliations among the 25 accessions based on the 22 core expressed sequence tag-simple sequence repeat (EST-SSR) markers; the 25 cultivars were separated into four groups using the similarity coefficient of 0.6 as the threshold

et al. 2012), ISSRs markers (Attari et al. 2016; Bublyk et al. 2021), and AFLP markers (Li et al. 2020) have been used in several studies on various iris species. In this study, 34 722 potential EST-SSRs were identified using transcriptome sequencing, which demonstrated that it was an effective method for mining a large number of EST-SSR markers in *I.* × *germanica*. The results showed that 21.4% of unigenes contained SSR loci, suggesting that SSRs were well distributed in the transcript sequence. Moreover, the 124 758 unigenes identified in this study can be used as references for further research on the physiology, biochemistry, and molecular genetics of *I.* × *germanica,* which lacks a reference genome. Previously, EST-SSR markers have been developed in other ornamental plant species, such as peony (Gilmore et al. 2013; Wu et al. 2014) and crape myrtle (Ye et al. 2019) based on transcriptome sequences.

The most abundant type of nucleotides repeat units was tri-nucleotide with five tandem repeats (6 880) followed by five tandem repeats (3 207) in *I.* × *germanica*. As frameshift mutations may disrupt normal reading frames and result in the insertion of the wrong amino acids into the protein sequence and/or the creation of a codon that stops the protein chain, the motifs except for the tri-nucleotide type may be restricted; consequently, tri-nucleotide repeats are the most abundant motifs for SSRs (Metzgar et al. 2000).

In this study, di- and tri-nucleotide motifs were two of the most frequent SSR motif types. Our results are also consistent with the reports of studies on SSR motif types in other plants, such as calla lily (Wei et al. 2016) and *Curcuma longa* (Annadurai et al. 2013). The most common di-nucleotide motifs type in *I.* × *germanica* was AG/CT (72.3%). In *I. brevicaulis* and *I. fulva,* EST-SSR markers were developed by constructing cDNA libraries, and 89.8% of these markers were the AG/CT di-nucleotide motif type (Lawson & Zhang 2006). AG/CT was the most abundant motif in the 5'UTR and ORF fractions in gymnosperms (Morgante et al. 2002). AG/CT repeats are highly abundant and have been widely targeted for EST-SSR marker development in plants because they are often highly polymorphic, more abundant in untranslated regions (UTRs) than coding sequences (CDSs), and can consistently amplify and yield robust SSR markers (Dieffenbach et al. 1993; Temnykh et al. 2001).

https://doi.org/10.17221/93/2024-CJGPB

In a study of *Syringa oblate*, the most abundant di-nucleotides motif was AT/TA (34.1%), as these motifs are carried by codons with high frequencies of Ile (AUA) and Tyr (UAU) (Yang et al. 2020).

In this study, tri-nucleotide motif was also dominant, and the most common tri-nucleotide motif was AGG/CCT, followed by AAG/CTT. In *C. alismatifolia*, AGG/CCT (13.82%) is the most abundant tri-nucleotide motif and is closely followed by CCG/CGG (12.94%) (Taheri et al. 2019). In *Triticum aestivum* (Yang et al. 2016) and *Panicum sumatrense* (Desai et al. 2021), tri-nucleotide repeats are also the most prevalent type. While in 15 grass genomes of Poaceae, the most abundant motifs are GA/TC, followed by monomer motifs A/T monomer and then GCG/CGC motif (Wang & Wang 2016). According to these results, the dominant type of motifs varied in different plant species. Among the polymorphic SSRs, 10 out of 22 were tri-nucleotide motif, which had five or more tandem repeats in this study. As smaller motifs are more likely to have genotyping errors due to slipped-strand mispairing (stuttering) during the polymerase chain reaction (PCR), longer SSR loci play more prominent allelic fluctuation (Taheri et al. 2018).

On the basis of LS and floret size, the American Iris Society has divided the bearded irises into six groups for garden judging (Moores 2001). Generally, the groups with longer flower stems also have larger florets but flower later than the groups with smaller florets. In this study, most floral traits were positively correlated with each other (Table 2), which is consistent with the group standards of the American Iris Society.

The results of this study showed that one floral trait was associated with multiple markers, and the same marker locus can be associated with multiple traits (Table 4). The marker IG1212F18 was associated with seven floral traits (FD, NF, LS, LFP, WSP, HF, and WF). Both traits, LS and HF, had a pretty high correlation with marker IG1212F18 with a *P*-value less than 0.0001. The Pearson correlation coefficient between traits LS and HF was 0.854, indicating that the genes controlling these two traits might be closely related. Since the markers mined in this study were based on the transcript sequence, these markers can be useful for detecting markers associated with specific traits in *I.* × *germanica*.

Genetic diversity and relationships of species in family Iridaceae have been evaluated using plastid markers (Ikinci et al. 2011), ISSR markers (Rasul et al. 2024), and MGE sequences (Bublyk et al. 2021). In studies on the genetic polymorphism of Iris pumila, ISSR markers were the most polymorphic among individuals, while markers of MGE sequences had the lowest polymorphism (Bublyk et al. 2021). The genetic relationships of 55 species of the Iris subgenus, *Scorpiris,* were revealed using six plastid markers, leading to the identification of novel and reasonable relationships between several species (Ikinci et al. 2011). In this study, EST-SSR markers were developed based on transcriptome sequencing data, which showed high polymorphism among cultivars and was used to evaluate the genetic relationship of these cultivars.

Autumn Circus (C3) and Autumn Tryst (C4) were estimated to have high similarity in the UPGMA dendrogram (Figure 2), as they have the same male parent, Earl of Essex, which is the progenitor of numerous re-blooming cultivars (Niswonger 2003). Both Autumn Circus and Autumn Tryst were reliable re-bloomers, of which the traits may be inherited from Earl of Essex. In *Iris loczyi*, all populations were grouped into three major clades in the UPGMA tree, which correlated with geographic distance (Zhang et al. 2021).

Further analysis was conducted among the once-bloomer and re-bloomer groups; however, there was no significant correlation within the group of re-bloomers or once-bloomers. This result may indicate that not all re-bloomers in this study were derived from the same ancestor. Their re-blooming traits may have been inherited from ancestors with different genetic backgrounds.

In the dendrogram, although the pink colour cultivars Chariots of Fire and Kaligazam, despite not being closely related in their pedigree, were clustered together. Niswonger (2003) reported that the pink traits appear when each of the four groups of chromosomes has a converting gene that converts carotene to lycopene. Thus, these pink cultivars might have similar genetic backgrounds, potentially having a homozygous genotype in the genes responsible for controlling pink traits.

As the white cultivars and cultivars of the plicata type (an iris that is 'stitched' or has stippled fall bands usually on a white or yellow ground) were both affected by anthocyanin inhibitor alleles (Spoon 2008), the white cultivar Immortality and the white plicata cultivar Girly Girl were clustered together in the UPGMA dendrogram (Figure 2). The white colour of Immortality was controlled by homozygous domi-

https://doi.org/10.17221/93/2024-CJGPB

nant anthocyanin inhibitor alleles, which repress the synthesis of anthocyanin pigments. On the other hand, the different dosage effects of anthocyanin inhibitor alleles might produce cultivars of the plicata type, such as Girly Girl, which has a white background and 1/4 inches of purple plicata markings.

CONCLUSION

In this study, 34 722 EST-SSR loci were identified from RNA sequencing data, which could be used to develop markers for *I. germamica* and other iris species. The genotypic relations of most cultivars evaluated by cluster tree were consistent with their pedigree-based relationships. Among the polymorphic primers screened from these EST-SSR loci, 11 EST-SSR markers were associated with floral traits, indicating that the EST-SSR loci identified in this study will be useful in marker-assisted selection breeding programs of *I. germamica*.

REFERENCES

Agarwal M., Shrivastava N., Padh H. (2008): Advances in molecular marker techniques and their applications in plant science. Plant Cell Report, 27: 617–631.

Annadurai R.S., Neethiraj R., Jayakumar V., Damodaran A.C., Rao A.S.N., Katta M.A., Gopinathan S., Sarma S.P., Senthilkumar V., Niranjan V., GopinathA., Mugasimangalam R.C. (2013): *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. PLoS ONE, 8: e56217.

Attari S.E., Shoor M., Neghab M.G., Tehranifar A., Shafaroudi S.M. (2016): Evaluation of genetic diversity of iris genotypes (*Iris* spp.) using ISSR. Journal of Horticultural Science, 30: 376–382.

Azimi M., Sadeghian S., Ahari V.R., Khazaei F., Hafashjani A.F. (2012): Genetic variation of Iranian *Iris* species using morphological characteristics and RAPD markers. International Journal of AgriScience, 2: 875–889

Bouck A., Vision T. (2010): The molecular ecologist's guide to expressed sequence tags. Molecular Ecology, 16: 907–924.

Brachi B., Faure N., Horton M., Flahauw E., Vazquez A., Nordborg M., Roux F. (2010): Linkage and association mapping of *Arabidopsis thaliana* lowering time in nature. PLoS Genetics, 6: e1000940.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007): TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23: 2633–2635.

Bublyk O., Parnikoza I., Kunakh V. (2021): Assessing the levels of polymorphism and differentiation in *Iris pumila* L. populations using three types of PCR markers. Cytology and Genetics, 55: 36–46.

Butiuc-Keul A., Coste A., Postolache D., Laslo V., Halmagyi A., Cristea V., Farkas A.A. (2022): Molecular characterization of prunus cultivars from Romania by microsatellite markers. Horticulturae, 8: 291.

Chavhan R.L., Sable S., Narwade A.V., Hinge V.R., Kalbande B.B., Mukherjee A.K., Chakrabarty P.K., Kadam U.S. (2023): Multiplex molecular marker-assisted analysis of significant pathogens of cotton (*Gossypium* sp.). Biocatalysis and Agricultural Biotechnology, 47: 102557.

Desai H., Hamid R., Ghorbanzadeh Z., Bhut N., Padhiyar S.M., Kheni J., Tomar R.S. (2021): Genic microsatellite marker characterization and development in little millet (*Panicum sumatrense*) using transcriptome sequencing. Scientific Reports, 11: 20620.

Dieffenbach C.W., Lowe T.M., Dveksler G.S. (1993): General concepts for PCR primer design. PCR Methods and Applications, 3: 30–37.

Doyle J.J., Doyle J.L. (1987): A rapid DNA isolation method for small quantities of fresh tissues. Phytochemical Bulletin, 19: 11–15.

Gilmore B., Bassil N., Nyberg A., Knaus B., Smith D., Barney D., Hummer K. (2013): Microsatellite marker development in Peony using Next Generation Sequencing. Journal of the American Society for Horticultural Science, 138: 64–74.

Guo Q., Guo L., Zhang L., Zhang L.X., Ma H., Guo D., Huo X. (2017): Construction of a genetic linkage map in tree peony (*paeonia* sect. *moutan*) using simple sequence repeat (SSR) markers. Scientia Horticulturae, 219: 294–301.

Hinge V.R., Shaikh I.M., Chavhan R.L., Deshmukh A.S., Shelake R.M., Ghuge S.A., Dethe A.M., Suprasanna P., Kadam U.S. (2022): Assessment of genetic diversity and volatile content of commercially grown banana (*Musa* spp.) cultivars. Scientific Reports, 12: 7979.

Ikinci N., Hall T., Lledó M.D., Clarkson J.J., Tillie N., Seisums A., Saito T., Harley M., Chase M.W. (2011): Molecular phylogenetics of the juno irises, *Iris* subgenus *Scorpiris* (Iridaceae), based on six plastid markers. Botanical Journal of the Linnean Society, 167: 281–300.

Jiang K., Gao H., Xu N., Tsang E.P.K., Chen X. (2011): A set of microsatellite primers for *Zostera japonica* (Zosteraceae). American Journal of Botany, 98: e236–e238.

Lawson M.J., Zhang L. (2006): Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biology, 7: R14.

Lee O.N., Park H.Y. (2017): Assessment of genetic diversity in cultivated radishes (*Raphanus sativus*) by agro-

https://doi.org/10.17221/93/2024-CJGPB

nomic traits and SSR markers. Scientia Horticulturae, 223: 19–30.

Li F., Sun Y., Liu C., Yuan Y., Zheng L., Chen X., Bao J. (2020): Genetic diversity and population structure in bearded iris cultivars derived from *Iris* × *germanica* L. and its related species *I. pumila* L., I. *variegata* L., *I. pallida* Lam. Genetic Resources and Crop Evolution, 67: 2161–2172.

Liu K., Muse S.V. (2005): Power Marker: An integrated analysis environment for genetic analysis. Bioinformatics, 21: 2128–129.

Liu Z., Zhang J., Wang Y., Wang H., Wang L., Zhang L., Xiong M., He W., Yang S., Chen Q., Chen T., Luo Y., Zhang Y., Tang H., Wang X. (2022): Development and cross-species transferability of novel genomic-SSR markers and their utility in hybrid identification and trait association analysis in Chinese Cherry. Horticulturae, 8: 222.

Mazzucato A., Papa R., Bitocchi E., Mosconi P., Nanni L., Negri V., Picarella M.E., Siligato F., Soressi G.P., Tiranti B., Veronesi F. (2008): Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. Theoretical and Applied Genetics, 116: 657–669.

Metzgar D., Bytof J., Wills C. (2000): Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Research, 10: 72–80.

Moores W (2001): Progenitors in modern reblooming irises. Bulletin of American Iris Society, 320: 28–30.

Morgante M., Hanafey M., Powell W. (2002): Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nature Genetics, 30: 194–200.

Niswonger D. (2003): Hybridizing tall bearded iris. Bulletin of American Iris Society, 330: 34–46.

Paun O., Schönswetter P. (2012): Amplified fragment length polymorphism: An invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. Methods in Molecular Biology, 862: 75–87.

Rasul K.S., Majeed H.O., Faraj J.M., Lateef D.D., Tahir M.A. (2024): Genetic diversity and relationships among *Iris aucheri* genotypes determined via ISSR and CDDP markers. Genetic Resources and Crop Evolution, 2024: 0925–9864.

Riangwong K., Wanchana S., Aesomnuk W., Saensuk C., Nubankoh P., Ruanjaichon V., Kraithong T., Toojinda T., Vanavichit A., Arikit S. (2020): Mining and validation of novel genotyping-by sequencing (GBS)-based simple sequence repeats (SSRs) and their application for the estimation of the genetic diversity and population structure of coconuts (*Cocos nucifera* L.) in Thailand. Horticulture Research, 7: 156.

Rohlf F.J. (2005): NTSYSpc: Numerical Taxonomy and Multivariate Analysis System. Version 2.2. New York, State University of New York.

Ronoh R., Linde M., Winkelmann T., Abukutsa-Onyango M., Fufa Dinssa F.F., Debener T. (2018): Development of next-generation sequencing (NGS)-based SSRs in African nightshades: Tools for analyzing genetic diversity for conservation and breeding. Scientia Horticulturae, 235: 152–159.

Santos C., Almeida N.P., Alves M.L., Horre R., Krezdorn N., Leitão S.T., Aznar-Fernández T., Rotter B., Winter P., Rubiales D., Patto M.C.V. (2018): First genetic linkage map of *Lathyrus cicera* based on RNA sequencing-derived markers: Key tool for genetic mapping of disease resistance. Horticulture Research, 5: 45.

Šarhanová P., Pfanzel S., Brandt R., Himmelbach A., Blattner F.R. (2018): SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. Ecology and Evolution, 8: 10817–10833.

Spoon D. (2008): Plicata locus color patterns in bearded irises. Bulletin of American Iris Society, 349: 95–107.

Sturtevant A.H. (1961): Iris genetics. Engineering and Science, 1961: 16–17.

Sturtevant A.H., Randolph L.F. (1945): Iris genetics. Bulletin of American Iris Society, 99: 54–66.

Sun M., Li M.R., Shi F., Li L., Liu Y., Li L.F., Xiao H. (2012): Genomic and EST-derived microsatellite markers for *Iris laevigata* (Iridaceae) and other congeneric species. American Journal of Botany, 99: e286-e288.

Taheri S., Abdullah T.L., Yusop M.R., Hanafi M.M., Sahebi M., Azizi P., Shamshiri R.R. (2018): Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules, 23: 399.

Taheri S., Abdullah T.L., Rafii M.Y., Harikrishna J.A., Werbrouck S.P.O., Teo C.H., Sahebi M., Azizi P. (2019): De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. Scientific Reports, 9: 3047.

Tang S., Okashah R.A., Cordonnier-Pratt M.M., Pratt L.H., Johnson V.E., Taylor C.A., Arnold M.L., Knapp S. (2009): EST and EST-SSR marker resources for Iris. BMC Plant Biology, 9: 72.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001): Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research, 11: 1441–1452.

The American Iris Society (2019): Iris Classification. Available at www.irises.org/ gardeners/care-classification/ classification/ (accessed 14 February 2024).

Thiel T., Michalek W., Varshney R., Graner A. (2003): Exploiting EST databases for the development and char-

https://doi.org/10.17221/93/2024-CJGPB

acterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, 106: 411–422.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. (2012): Primer3 – new capabilities and interfaces. Nucleic Acids Research, 40: e115.

Upadhyay A., Kadam U.S., Priya C.M., Karibasappa G.S. (2010a): Microsatellite and RAPD analysis of grape (*Vitis* spp.) accessions and identification of duplicates/misnomers in germplasm collection. Indian Journal of Horticulture, 67: 8.

Upadhyay A., Kadam U.S., Priya C.M., Karibasappa G.S. (2010b): Microsatellite analysis to differentiate clones of Thompson seedless grapevine. Indian Journal of Horticulture, 67: 260–263.

Varshney R.K., Graner A., Sorrells M.E. (2005): Genetic microsatellite markers in plants: Features and applications. Trends in Biotechnology, 23: 48–55.

Wang X., Wang L. (2016): GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. Frontiers in Plant Science, 7: 1350.

Wang Y., Jia H., Shen Y., Zhao H., Yang Q., Zhu C., Sun D., Wang C., Zhou C., Jiao Y., Chai C., Yan L., Li X., Jia H., Gao Z. (2020): Construction of an anchoring SSR marker genetic linkage map and detection of a sex-linked region in two dioecious populations of red bayberry. Horticulture Research, 7: 53.

Wei Z., Sun Z., Cui B., Zhang Q., Xiong M., Wang X., Zhou D. (2016): Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: *De novo* assembly, annotation and EST-SSR marker development. Peer J, 4: e2378.

Williams R.C. (1989): Restriction fragment length polymorphism (RFLP). American Journal of Physical Anthropology, 32: 159–184.

Wu J., Cai C., Cheng F., Cui H., Zhou H. (2014): Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. Molecular Breeding, 34: 1853–1866.

Xiao Y., Hu Y., Liu M., Chen X. (2012): Isolation and characterization of polymorphic microsatellites in *Iris ensata* (Iridaceae). American Journal of Botany, 99: e498–e500.

Yang Y., He R., Zheng J., Hu Z., Wu J., Leng P. (2020): Development of EST-SSR markers and association mapping with floral traits in *Syringa oblata*. BMC Plant Biology, 20: 436.

Yang Z., Peng Z., Yang H. (2016): Identification of novel and useful EST-SSR markers from de novo transcriptome sequence of wheat (*Triticum aestivum* L.). Genetics and Molecular Research, 15: 15017509.

Ye Y., Feng L., Liang X. (2019): Characterization, validation, and cross-species transferability of newly developed EST-SSR markers and their application for genetic evaluation in crape myrtle (*Lagerstroemia* spp). Molecular Breeding, 39: 26.

Yu J., Buckler E.S. (2006): Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17: 155–160.

Zhang G., Han Y., Wang H., Wang Z., Xiao H., Sun M. (2021): Phylogeography of *Iris loczyi* (Iridaceae) in Qinghai-Tibet Plateau revealed by chloroplast DNA and microsatellite markers. AoB PLANTS, 13: plab070.

Zhen Q., Fang T., Peng Q., Zhao L., Owiti A. (2018): Developing gene-tagged molecular markers for evaluation of genetic association of apple *SWEET* genes with fruit sugar accumulation. Horticulture Research, 5: 14.

Zhu J., Zhang J., Jiang M., Wang W., Jiang J., Li Y., Yang L. Zhou X. (2021): Development of genome-wide SSR markers in rapeseed by next generation sequencing. Gene, 798: 145798.

Received: August 6, 2024
Accepted: December 6, 2024
Published online: February 3, 2025